DOCUMENT RESUME

ED 304 157 IR 052 666

AUTHOR Ajiferuke, Isola
TITLE A Probabilistic Model for the Distribution of Authorships: A Preliminary Report.
PUB DATE 20 May 88
NOTE 25p.; Paper presented at the American Society of Information Science Student Research Mini Conference (Syracuse, NY, May 20, 1988).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Abstracting; *Authors; *Bibliometrics; *Goodness of Fit; *Mathematical Models; *Periodicals; *Probability; Social Influences; Statistical Analysis
IDENTIFIERS Collaborative Research; Gaussian Poisson Distribution; University of Western Ontario (Canada)

ABSTRACT

The purpose of this study was to develop a model for the distribution of authorships--based on the initial hypothesis that the distribution of authorships follows a shifted Waring distribution--and to test the derived model and some other discrete probability models for goodness-of-fit against empirical data. Bibliographic data from 15 abstracting journals covering the literature in six fields--engineering, medical, physical, mathematical and social sciences, and humanities--were used in testing the goodness-of-fit of the shifted Waring distribution and 13 other discrete probability models. The preliminary findings presented here are based on 60 data sets collected from 10 abstracting journals covering the literature in the mathematical and social sciences and humanities. They indicate that the promising models for the distribution of authorships are the shifted Waring, shifted generalized negative binomial, shifted negative binomial, shifted generalized Poisson, and shifted inverse Gaussian-Poisson distributions. Three advantages and possible practical applications of a model for the distribution of authorships include: (1) ability to summarize the entire frequency distribution by a few parameters of the model; (2) estimation of the number of entries in an author index; and (3) usefulness in a simulation study designed to determine, subject to space constraints, the maximum number of authors per paper to be included in an author index. Forms of the discrete probability models are appended. (5 tables, 12 references) (Author/CGD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

A PROBABILISTIC MODEL FOR THE DISTRIBUTION OF AUTHORSHIPS : A PRELIMINARY REPORT

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

Minor changes have been made to improve reproduction quality.

Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

by

Isola Ajiferuke
School of Library and Information Science
University of Western Ontario, London,
Ontario, Canada N6G 1H1

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Isola Ajiferuke
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ABSTRACT

The purpose of the study is to develop a model for the distribution of authorships. The initial hypothesis is that the distribution of authorships follows a shifted Waring distribution. Bibliographic data from 15 abstracting journals covering the literature in 6 fields (engineering sciences, medical sciences, physical sciences, mathematical sciences, social sciences, and humanities) are to be used in testing the goodness-of-fit of the shifted Waring distribution and 13 other discrete probability models. The findings presented here are based on data collected from only 10 abstracting journals covering the literature in the mathematical sciences, social sciences, and humanities. The shifted Waring, shifted generalized negative binomial, shifted negative binomial, shifted generalized Poisson and shifted generalized Gaussian-Poisson distributions performed fairly well in the Chi Square tests.

BEST COPY AVAILABLE

ED304157
IR052666

1. INTRODUCTION

Bibliometricians have developed models for various social phenomena. The three most common models are Bradford's law (1934) for journal productivity, Lotka's law (1926) for scientific productivity, and Zipf's law (1949) for frequency of words in text. Models that have been established for other social phenomena are listed in articles by Simon (1955), Kendall (1961), Haitun (1982a), and Chen and Leimkuhler (1986).

One social phenomenon for which no model has been developed is collaboration in research, though a few have been suggested without their goodness-of-fit being tested. Just as the number of papers published in scholarly journals is taken as an indication of productivity, co-authorships of papers in journal is often taken as an indication of collaboration in research.

Price and Beaver (1966), using memos circulated among members of an invisible college, inferred a Poisson model for the distribution of authorships while Haitun (1982b), using the same data, classified the distribution of authorships among those stationary scientometric distributions that cannot be approximated by the Zipf distribution.

Worthen (1978) observed that for drug literatures the number of papers with one author makes up about 1/3 of the literature, those

with two authors about 1/3 of the remaining literature, those with three, 1/3 of the remainder, and so on. Goffman and Warren (1980), on the other hand, observed that for the schistosomiasis literature, the number of publications with one author makes up about 2/3 of the literature, those with two authors about 2/3 of the remaining literature, those with three, 2/3 of the remainder, and so on. The two observations implied a geometric model for the distribution of authorships.

The objectives of this research, therefore, are to derive a theoretical model for the distribution of authorships and to test the derived model and some other discrete probability models (for goodness-of-fit) against empirical data .

2. PROBABILITY MODELS

2.1 Theoretical Model

The initial theoretical model being proposed is the shifted Waring distribution. Its derivation is as follows :

Assume that a researcher completes a project alone if he can. However, if he cannot, he brings in additional researchers, one at a time, until the project is completed. Let an attempt by the researcher(s) to complete a project without bringing in an additional researcher represent a trial. The probability that the project will be completed on any trial is constant. The distribution of the number of trials or the number of researchers required before the project is completed can then be described by the geometric

distribution. However, the probability of completing the project at any trial without bringing in an additional researcher, i.e., p, would vary from project to project (types of projects include review, opinion on a topic, experimental research, theoretical work, etc). Thus,

$$P(X = r \,/p) = p\,(1-p)^{r-1} \quad ; \; r = 1,2,\ldots$$

For analytical convenience, we assume that p varies as a beta distribution with parameters α,β, i.e., p has a density function

$$f(p) = \frac{1}{B(\alpha,\beta)}\, p^{\alpha-1}\,(1-p)^{\beta-1} \qquad \text{if } 0 \le p \le 1$$

$$0 \qquad \text{otherwise}$$

where $\alpha > 0$, $\beta > 0$ are constants and

$$B(\alpha,\beta) = \int_0^1 x^{\alpha-1}\,(1-x)^{\beta-1}\,dx$$

$$= \frac{\Gamma(\alpha)\,\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

Hence,

$$P(X = r) = E\,[P(X = r\,/p)]$$

$$= \int P(X = r\,/p)\, f(p)\, dp$$

$$= \int_0^1 p(1-p)^{r-1}\, \frac{p^{\alpha-1}\,(1-p)^{\beta-1}}{B(\alpha,\beta)}\, dp$$

$$= \frac{1}{B(\alpha,\beta)} \int_0^1 p^{\alpha}\,(1-p)^{r+\beta-2}\, dp$$

$$= \frac{1}{B(\alpha,\beta)} \; B(\alpha+1, r+\beta-1)$$

$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)} \; \frac{\Gamma(\alpha+1)\,\Gamma(r+\beta-1)}{\Gamma(\alpha+\beta+r)}$$

$$= \frac{\alpha}{\alpha+\beta} \; \frac{\Gamma(\alpha+\beta+1)}{\Gamma(\alpha+\beta+r)} \; \frac{\Gamma(r+\beta-1)}{\Gamma(\beta)}$$

$$= \frac{\alpha}{\alpha+\beta} \prod_{k=2}^{r} \frac{(k+\beta-2)}{(\alpha+\beta+k-1)} \quad ; \; \alpha > 0, \; \beta > 0, \; r = 1,2,\ldots$$

This distribution is known as the shifted Waring.

2.2 Empirical Models

In addition to the shifted Waring distribution, other discrete distributions that will be tested are :

1. Zipf
2. Mandelbrot
3. Geometric
4. Shifted Poisson
5. Shifted Generalized Poisson
6. Logarithmic
7. Borel-Tanner
8. Shifted Yule
9. Shifted Generalized Waring
10. Shifted Inverse Gaussian-Poisson (IGP)
11. Shifted Binomial
12. Shifted Negative Binomial
13. Shifted Generalized Negative Binomial

The form of each model and the method of estimation of its

parameters are given in the Appendix.

The Chi Square test will be used to determine the goodness-of-fit of the models at 5% and 1% levels of significance. The Chi Square test was chosen in preference to the Kolmogorov-Smirnov test because the latter is strictly applicable only if the data are continuously distributed (Conover, 1980).

3. DATA COLLECTION

Most bibliometric studies of research collaboration have used either scholarly journals or abstracting journals or both as their sources of data. Because of their comprehensive nature, abstracting journals will be used as the source of data for this study. However, only research papers published in scholarly journals or presented at conferences will be counted. Books, dissertations, theses, and research papers by anonymous or corporate authors are being excluded because :

(1) Some books are made up of several articles, which are at times written by different authors, and not all the abstracting journals cover monographic literature. However, book chapters or parts of books abstracted separately will be counted.

(2) Dissertations and theses are of single authorships by nature.

(3) One cannot easily ascertain the number of people responsible for research works credited to anonymous or corporate authors. Also,

editors, compilers, translators, and moderators of symposia will not be regarded as co-authors since they are not intellectually responsible for the research works attributed to them.

In order to see how well the probability models fit the distribution of authorships in various fields, data will be collected from engineering sciences, medical sciences, physical sciences, mathematical sciences, social sciences, and humanities.

Engineering Index abstracts literature from all the disciplines in engineering sciences while **Index Medicus** does the same for the disciplines in medical sciences. However, no single abstracting journal covers all the disciplines in either the physical sciences, mathematical sciences, social sciences, or humanities. Hence, a few abstracting journals have been chosen to represent each of these fields (papers from the abstracting journals representing each field will not be added together because of journal overlap). Availability, scope, and prominence are the factors taken into consideration in choosing the abstracting journals. The list of the chosen abstracting journals and the range of the number of records abstracted per year by each journal are given below in Table 1.

Six annual cumulative issues of each abstracting journal will be used in the study. Because some of the chosen abstracting journals came into being late in the fifties or early in the sixties, the years chosen by systematic sampling are 1961, 1966, 1971, 1976, 1981, and 1986. However, data for philosophy will be collected for 1967/68, 1969, 1971, 1976, 1981 and 1986 because the first issue of Philosopher's Index was published in 1967/68.

For each of **Engineering Index, Index Medicus, Chemical Abstracts, Biological Abstracts, Physics Abstracts** and **Psychological Abstracts**, a few issues will be selected at random for each year. For the others, which index relatively small number of records per year, all the issues will be used.

Articles contained in each annual cumulative issue of each abstracting journal will be sorted into different levels of authorships. The number of articles at each level will then be counted and the resulting distribution of authorships taken as one data set. Hence, there will be a total of 90 data sets to be used in the analysis.

Table 1. List of the abstracting journals selected and the range of the number of records abstracted per year by each journal

Abstracting Journal	Field	Number of Records
1. Engineering Index (1917-	Engineering Sciences	37000 - 131000
2. Index Medicus (1960-	Medical Sciences	110000 - 400000
3. Physics Abstracts (1903-	Physical Sciences	21000 - 130000
4. Chemical Abstracts (1924-	Physical Sciences	145000 - 240000
5. Biological Abstracts (1926/27-	Physical Sciences	87000 - 235000
6. Mathematical Reviews (1940-	Mathematical Sciences	12000 - 40000
7. Statistical Theory & Methodology Abstracts (1960-	Mathematical Sciences	800 - 4000
8. Computer Abstracts (1957-	Mathematical Sciences	3000 - 4000
9. Sociological Abstracts (1953-	Social Sciences	2000 - 11000
10. Inter. Pol. Sc. Abstracts (1951-	Social Sciences	1500 - 6500
11. Economics Abstracts (1953-	Social Sciences	2000 - 7000
12. Psychological Abstracts (1922-	Social Sciences	14000 - 35000
13. Historical Abstracts (1955-	Humanities	3000 - 23000
14. Abstracts of English Studies (1958-	Humanities	2500 - 4000
15. Philosopher's Index (1967/68-	Humanities	2000 - 7000

4. RESULTS OF THE PILOT STUDY.

At present, data have been collected from only 10 abstracting journals covering the literature in the mathematical sciences, social sciences, and humanities. Hence, the findings presented below are based on only 60 data sets.

In the humanities, many of the models performed very well with shifted generalized Poisson and shifted inverse Gaussian-Poisson doing exceptionally well (see Table 2).

In the social sciences, only shifted generalized Poisson, shifted generalized inverse Gaussian-Poisson and shifted negative binomial distributions did very well while shifted Waring and shifted generalized negative binomial distributions performed fairly well (see Table 3).

In the mathematical sciences, all the models performed poorly (see Table 4).

Overall, only shifted generalized Poisson, shifted Waring, shifted generalized inverse Gaussian-Poisson, shifted negative binomial and shifted generalized negative binomial distributions performed fairly well in the Chi Square tests (see Table 5).

Table 2 : Results of the Chi Square goodness-of-fit tests for the Humanities (18 data sets)

Model	No. of Valid Results	No. of Passes 5%	No. of Passes 1%	No. of Best Fits Provided
Zipf	18	9	12	1
Mandelbrot	18	14	16	4
Geometric	18	6	10	0
Shifted Poisson	18	7	7	1
Sh. Gen. Poisson	18	17	17	1
Logarithmic	18	7	7	0
Borel-Tanner	18	7	9	0
Shifted Yule	18	10	13	3
Shifted Waring	17	14	17	0
Sh. Gen. Waring	3	0	0	0
Shifted IGP	17	16	17	6
Shifted Binomial	1	1	1	0
Sh. Neg. Binomial	17	14	16	0
Sh. Gen. Neg. Bin.	14	13	14	2

Note : 1) "No. of valid results" refers to the number of data sets in which the assumptions of the model are not violated.

2) "No. of passes" refers to the number of times that there was no significant difference between the observed distribution and the expected distribution at the specified level of significance.

3) "No. of best fits provided" refers to the number of times that the model had the highest probability level for the Chi Square value

Table 3 : Results of the Chi Square goodness-of-fit tests for the Social Sciences (24 data sets)

Model	No. of Valid Results	No. of Passes 5%	No. of Passes 1%	No. of Best Fits Provided
Zipf	24	2	2	0
Mandelbrot	24	5	6	0
Geometric	24	5	7	0
Shifted Poisson	24	3	4	1
Sh. Gen. Poisson	24	18	22	2
Logarithmic	24	6	8	0
Borel-Tanner	24	6	7	0
Shifted Yule	24	5	5	1
Shifted Waring	12	10	12	2
Sh. Gen. Waring	13	1	2	0
Shifted IGP	24	19	24	13
Shifted Binomial	0	0	0	0
Sh. Neg. Binomial	24	17	20	3
Sh. Gen. Neg. Bin.	14	10	13	2

Table 4 : Results of the Chi Square goodness-of-fit tests for the Mathematical Sciences (18 data sets)

Model	No. of Valid Results	No. of Passes 5%	No. of Passes 1%	No. of Best Fits Provided
Zipf	18	0	0	0
Mandelbrot	18	0	0	0
Geometric	18	4	5	1
Shifted Poisson	18	2	4	0
Sh. Gen. Poisson	18	7	9	2
Logarithmic	18	0	0	0
Borel-Tanner	18	0	0	0
Shifted Yule	18	0	0	0
Shifted Waring	0	0	0	0
Sh. Gen. Waring	4	0	0	0
Shifted IGP	14	5	6	13
Shifted Binomial	4	3	3	0
Sh. Neg. Binomial	14	3	5	0
Sh. Gen. Neg. Bin.	5	4	5	2

Table 5 : Results of the Chi Square goodness-of-fit tests for the three fields combined together (60 data sets)

Model	No. of Valid Results	No. of Passes 5%	No. of Passes 1%	No. of Best Fits Provided
Zipf	60	11	14	1
Mandelbrot	60	19	22	4
Geometric	60	15	22	1
Shifted Poisson	60	12	15	2
Sh. Gen. Poisson	60	42	48	5
Logarithmic	60	13	15	0
Borel-Tanner	60	13	16	0
Shifted Yule	60	15	18	4
Shifted Waring	29	24	29	2
Sh. Gen. Waring	20	1	2	0
Shifted IGP	55	40	47	32
Shifted Binomial	5	4	4	0
Sh. Neg. Binomial	55	34	41	3
Sh. Gen. Neg. Bin.	33	27	32	6

5. CONCLUSION

Judging from the analysis of the 60 data sets collected from 10 abstracting journals covering the literature in the mathematical sciences, social sciences, and humanities, the promising models for the distribution of authorships are the shifted generalized Poisson, shifted Waring (the proposed theoretical model), shifted inverse Gaussian-Poisson, shifted negative binomial and shifted generalized negative binomial distributions. However, we have to wait for the results of the analysis of the remaining 30 data sets, which are to be collected from 5 abstracting journals covering the literature in the engineering sciences, medical sciences and physical sciences, before determining the best model for the distribution of authorships.

The advantages and possible practical applications of a model for the distribution of authorships include :

(i) Ability to summarize the entire frequency distribution by a few parameters of the model.

(ii) Estimation of the number of entries in an author index : Abstracting services often add extra entries in case of multiple authorships. For large abstracting journals and indexes, it would be time consuming to determine ahead of publication how many author entries would be made for N, a fixed number, papers. So, in order to estimate the number of entries, an abstractor may randomly sample n out of the N papers and estimate the parameters of the model from

the sample. The values of the parameters could then be used to estimate for the N papers, the number of papers with one author, two authors, three authors, and so on. The abstractor then should be able to estimate the total number of entries, and, hence, have a rough estimate of the size of the author index.

(iii) Usefulness in a simulation study designed to determine, subject to space constraints, the maximum number of authors per paper to be included in an author index. Only the first author would be included for papers with more than the maximum number of authors per paper allowed.

APPENDIX : FORMS OF THE DISCRETE PROBABILITY MODELS

In the following, we write

$f(x)$ = frequency of occurrence of x

n = sample size

MLE = maximum likelihood estimate

$\bar{x}$ = sample mean

s^2 = sample variance

1. Zipf

$$p(x) = k\, x^{-b}$$

where $0 < k < 1$, $b > 0$ for $x = 1,2,....$

The maximum likelihood estimator of b satisfies the equation

$$f(x)\ \ln x \,/\, \Sigma f(x) = -\zeta'(\hat{b}) \,/\, \zeta(b)\ ,$$

where $\zeta(.)$ denotes the Riemann zeta function ; and

$$k = [\Sigma\, x^{-b}\ ]^{-1}$$

2. Mandelbrot

$$p(x) = \frac{k}{n}\,(x+c)^{-b}\ ;$$

where $k > 0$, $b > 0$, $-1 < c < \infty$ for $x = 1,2,...$

The parameters are estimated by non-linear least-squares regression method.

3. Geometric

$$p(x) = p\,(1-p)^{x-1}\ ;$$

where $0 < p < 1$ for $x = 1,2,...$

The MLE of p is $\hat{p} = 1/\bar{x}$

4. Shifted Poisson

$$p(x) = \frac{e^{-\lambda}\lambda^{x-1}}{(x-1)!} \quad ;$$

where $\lambda > 0$ for $x = 1,2,\ldots$

MLE of λ is $\hat{\lambda} = \bar{x} - 1$

5. Shifted Generalized Poisson

$$p(x) = \frac{\lambda_1(\lambda_1 + (x-1)\lambda_2)^{x-2} e^{-(\lambda_1 + (x-1)\lambda_2)}}{(x-1)!} \quad ;$$

where $\quad > 0, \ | \ | < 1$ for $x = 1,2,\ldots$

The moment estimates of λ_1 and λ_2 are

$$\hat{\lambda}_2 = 1 - ((\bar{x}-1)/s^2)^{\frac{1}{2}} \quad \text{and}$$

$$\hat{\lambda}_1 = (\bar{x}-1)(1 - \hat{\lambda}_2)$$

The maximum likelihood estimator of λ_2 is given by the root of the equation

$$\sum_{x=1}^{x_{max}} \frac{f(x)(x-1)(x-2)}{(\bar{x}-1 + (x-\bar{x})\hat{\lambda}_2)} - n(\bar{x}-1) = 0$$

and

$$\hat{\lambda}_1 = (\bar{x}-1)(1 - \hat{\lambda}_2)$$

6. Logarithmic Series

$$p(x) = \frac{\alpha\theta^x}{x} \quad ;$$

where $0 < \theta < 1$, $\alpha = -[\ln(1-\theta)]^{-1}$ for $x = 1,2,\ldots$

The moment estimate of θ is

$$\hat{\theta} = 1 - \frac{\bar{x}}{(s^2 + \bar{x}^2)}$$

The maximum likelihood estimator of θ is given by the root of the equation

$$\bar{x} + \frac{\hat{\theta}}{(1 - \hat{\theta}) \ln (1 - \hat{\theta})} = 0 ;$$

7. Borel - Tanner

$$p(x) = \frac{x^{x-2} \alpha^{x-1} e^{-\alpha x}}{(x-1)!} ;$$

where $\alpha > 0$ for $x = 1,2,...$

The moment estimate of α is given by

$$\hat{\alpha} = 1 - 1/\bar{x}$$

8. Shifted Yule

$$p(x) = \frac{\alpha \Gamma(x) \Gamma(\alpha+1)}{\Gamma(\alpha+x+1)} ;$$

where $\alpha > 0$ for $x = 1,2,...$

The moment estimate of α is given by

$$\hat{\alpha} = 1 + 1/(\bar{x}-1) = \bar{x}/(\bar{x}-1)$$

9. Shifted Waring

$$p(x) = \frac{\alpha}{\alpha + \beta} \frac{\Gamma(\alpha+\beta+1) \Gamma(x+\beta-1)}{\Gamma(\alpha+\beta+x)\Gamma(\beta)} ;$$

where $\alpha > 0$, $\beta > 0$ for $x = 1,2,...$

The moment estimates of α and β are given by

$$\hat{\alpha} = \frac{2s^2}{(s^2 - \bar{x}(\bar{x}-1))} ; \text{ and}$$

$$\hat{\beta} = (\hat{\alpha}-1)(\bar{x}-1)$$

Note that the mean exists only if $\alpha > 1$ while the variance

exists only if $\alpha > 2$.

The maximum likelihood estimator of α is given by the root of the equation

$$\frac{n(\bar{x}-1)(\hat{\alpha}-1)}{\hat{\alpha}[\bar{x}(\hat{\alpha}-1)+1]} - \sum_{k=2}^{x_{max}} f(k) \sum_{i=2}^{k} \frac{1}{[x(\alpha-1)+x]} = 0 \; ; \text{ and}$$

$$\hat{\beta} = (\hat{\alpha}-1)(\bar{x}-1)$$

10. Shifted Generalized Waring

$$p(x) = \frac{\Gamma(v+\alpha)\,\Gamma(x+v-1)\,\Gamma(x+\beta-1)}{B(\alpha,\beta)\,\Gamma(v)\,\Gamma(x+v+\alpha+\beta-1)\,(x-1)!} \; ;$$

where $\alpha > 0$, $\beta > 0$, $v > 0$ for $x = 1,2,\ldots$

The three parameters can be estmated from the first three ascending factorial moments given by

$$\mu_{[1]} = \frac{\beta v}{\alpha-1} \; ; \quad \alpha > 1$$

$$\mu_{[2]} = \frac{\beta v\,(v+1)(\beta+1)}{(\alpha-1)(\alpha-2)} \; ; \quad \alpha > 2 \text{ and}$$

$$\mu_{[3]} = \frac{\beta v\,(v+1)(v+2)(\beta+1)(\beta+2)}{(\alpha-1)(\alpha-2)(\alpha-3)} \; ; \quad \alpha > 3$$

11. Shifted Inverse Gaussian-Poisson

$$p(x) = (2\alpha/\pi)^{\frac{1}{2}}\, e^{\alpha(1-\theta)^{\frac{1}{2}}} \frac{(\frac{1}{2}\alpha\theta)^{x-1}}{(x-1)!}\, K_{x-3/2}(\alpha) \; ;$$

where $0 \le \theta \le 1$, $\alpha \ge 0$, $K_v(z)$ is the modified Bessel function of the second kind of order v with argument z for $x = 1,2,\ldots$

Estimation of parameters :

(a) If the observed frequency distribution is reversed J-shape, with a large proportionate frequency in

the unit cell.

$$\hat{\theta} = 1 - \left[\frac{-\ln p(x=1)}{2(\bar{x}-1) + \ln p(x=1)}\right]^2 \quad ; \text{ and}$$

$$\hat{\alpha} = \frac{2(\bar{x}-1)}{\hat{\theta}} (1 - \hat{\theta})^{\frac{1}{2}}$$

(b) If observed frequency distribution is unimodal.

The moment estimates of θ and α are given by

$$\hat{\theta} = 1 - \left[\frac{2s^2}{(\bar{x}-1)} - 1\right]^{-1} \quad ; \text{ and}$$

$$\hat{\alpha} = \frac{2(\bar{x}-1)}{\hat{\theta}} (1 - \hat{\theta})^{\frac{1}{2}}$$

The maximum likelihood estimator of α is given by the root of the equation

$$[\hat{\alpha}(\epsilon^2+\hat{\alpha}^2)^{-\frac{1}{2}}][1 + (\bar{x}-1)/w] - (1/n) \sum_{x=1}^{x_{max}} f(x) \, R_{x-3/2}(\hat{\alpha}) = 0$$

where $\epsilon = \bar{x} - 1$, $w = (\epsilon^2+\hat{\alpha}^2)^{\frac{1}{2}} - \epsilon$, $R_v(z) = \frac{K_{v+1}(z)}{K_v(z)}$

and

$$\hat{\theta} = \frac{-2\epsilon^2 \pm 2\epsilon(\epsilon^2+\hat{\alpha}^2)^{\frac{1}{2}}}{\hat{\alpha}^2}$$

12. Shifted Binomial

$$p(x) = \binom{v}{x-1} p^{x-1} (1-p)^{v-(x-1)} \quad ;$$

where $0 < p < 1$, $v > 0$ for $x = 1,2, ..,v,v+1$

The moment estimates of p and v are given by

$$\hat{p} = 1 - \frac{s^2}{\bar{x}-1} \quad ; \text{ and}$$

$$\hat{v} = \frac{\bar{x}-1}{\hat{p}}$$

The maximum likelihood estimator of v is given by the root of the equation

$$\sum_{j=1}^{x_{max}-1} (\hat{v}-j+1)^{-1} f_j + n \ln [1 - (\bar{x}-1)/\hat{v}] = 0 ;$$

where f_j = number of x's which exceed j ; and

$$\hat{p} = \frac{\bar{x}-1}{\hat{v}}$$

13. Shifted Negative Binomial

$$p(x) = \binom{v+x-2}{x-1} p^v (1-p)^{x-1} ;$$

where $0 < p < 1$, $v > 0$ for $x = 1,2,...$

The momemt estimates of p and v are given by

$$\hat{p} = \frac{\bar{x}-1}{s^2} \quad ; \text{ and}$$

$$\hat{v} = \frac{\hat{p}(\bar{x}-1)}{1-\hat{p}}$$

The maximum likelihood estimator of v is given by the root of the equation

$$\ln [1 + (\bar{x}-1)/\hat{v}] - \sum_{j=2}^{x_{max}} \frac{F_j}{(\hat{v}+j-2)} = 0 ,$$

where Fj = proportion of x's which are greater than or equal to j; and

$$\hat{p} = \frac{\hat{v}}{(\bar{x}+\hat{v}-1)}$$

14. Shifted Generalized Negative Binomial

$$p(x) = \frac{v\ \Gamma[v+\beta(x-1)]}{(x-1)!\ \Gamma[v+\beta(x-1)-(x-1)+1]} \alpha^{x-1} (1-\alpha)^{v+\beta(x-1)-(x-1)} ;$$

where $0 < \alpha < 1$, $|\alpha\beta| < 1$, $v > 0$ for $x = 1,2, \ldots$

The moment esimates of α, β, and v are given by

$$\hat{\alpha} = 1 - \tfrac{1}{2}A + (\tfrac{1}{4}A^2 - 1)^{\frac{1}{2}} ;$$

$$\hat{\beta} = 1/\hat{\alpha}\,[1 - ((\bar{x}-1)(1-\hat{\alpha})/\mu_2)^{\frac{1}{2}}] \text{ ; and}$$

$$\hat{v} = (\bar{x}-1)(1-\hat{\alpha}\hat{\beta})/\hat{\alpha}$$

where $$A = -2 + \frac{[(\bar{x}-1)\mu_3 - 3\mu_2^2]^2}{(\bar{x}-1)\mu_2^3}$$

and μ_2 and μ_3 are the second and third central moments respectively.

REFERENCES

Bradford, S.C. (1934). Sources of information on specific subjects. **Engineering**, 137, 85-96.

Chen, Y. & Leimkuhler, F.F. (1986). A relationship between Lotka's law, Bradford's law, and Zipf's law. **Journal of the American Society for Information Science**, 37(5), 307-314.

Conover, W.J. (1980). **Practical nonparametric statistics**. 2nd edition. New York : John Wiley & Sons.

Goffman, W. & Warren, K.S. (1980). **Scientific information systems and the principle of selectivity**. New York : Praegler Publishers.

Haitun, S.D. (1982a). Stationary scientometric distributions : Part I. Different approximations. **Scientometrics**, 4(1), 5-25.

Haitun, S.D. (1982b). Stationary scientometric distributions : Part II. Non-Gaussian nature of scientific activities. **Scientometrics**, 4(2), 89-104.

Kendall, M.G. (1961). Natural law in the social sciences. **Journal of the Royal Statistical Society, Series A**, 124(Part 1), 1-16.

Lotka, A.J. (1926). The frequency distribution of scientific productivity. **Journal of the Washington Academy of Science**, 16(12), 317-323.

Price, D.J. DE S. & Beaver, D. deB. (1966). Collaboration in an invisible college. **American Psychologist**, 21, 1011-1018.

Simon, H.A. (1955). On a class of skew distribution functions. **Biometrika**, 42(3/4), 425-440.

Worthen, D.B. (1978). Short-lived technical literatures : A bibliometric analysis. **Methods of Information in Medicine**, 17(3), 190-198.

Zipf, G.K. (1949). **Human behavior and the principle of least effort**. New York : Addison-Wesley Press, Inc.